AFRL-HE-WP-TP-2006-0096

Informational and Energetic Masking Effects in Multitalker Speech Perception

Douglas S. Brungart

AFRL/HECB

Interim Report

20061102008

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Battlespace Acoustics Branch
Wright-Patterson AFB OH 45503-7901

REPORT DOCUMENTATION PAGE			*Form Approved* OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	1 Oct 2006
2. REPORT TYPE	Interim
3. DATES COVERED *(From - To)*	
4. TITLE AND SUBTITLE	Informational and Energetic Masking Effects in Multitalker Speech
5a. CONTRACT NUMBER	
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	61102F
6. AUTHOR(S)	Douglas S. Brungart
5d. PROJECT NUMBER	2313
5e. TASK NUMBER	HC
5f. WORK UNIT NUMBER	52
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	
8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	Air Force Materiel Command Air Force Research Laboratory Human Effectiveness Directorate Warfighter Inteface Division Battlespace Acoustics Branch Wright-Patterson AFB OH 45433-7901
10. SPONSOR/MONITOR'S ACRONYM(S)	AFRL/HECP
11. SPONSORING/MONITORING AGENCY REPORT NUMBER	AFRL-HE-WP-TP-2006-0096
12. DISTRIBUTION AVAILABILITY STATEMENT	Approved for public release; distribution is unlimited.
13. SUPPLEMENTARY NOTES	Cleared as ASC-04-0918, Apr 02, 2004.
14. ABSTRACT	
15. SUBJECT TERMS	

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
UNCL	UNCL	UNCL	SAR	5	Douglas S. Brungart
					19b. TELEPONE NUMBER *(Include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI-Std Z39-18

Informational and Energetic Masking Effects in Multitalker Speech Perception

Douglas S. Brungart
Air Force Research Laboratory

Introduction:

When a speech signal is obscured by a second simultaneous competing speech signal, two types of masking contribute to overall performance. Traditional "energetic" masking occurs when both utterances contain energy in the same critical bands at the same time and portions of one or both of the speech signals are rendered inaudible at the periphery. Higher-level "informational masking" occurs when the signal and masker are both audible but the listener is unable to disentangle the elements of the target signal from a similar-sounding distracter. Because "informational masking" is restricted to cases where the masking signal is similar to the target signal, it has a much greater impact on performance when a speech signal is masked by speech than it does when a speech signal is masked by noise. Furthermore, its effects depend specifically on the characteristics of the target and masking speech signals. This brief chapter outlines the results of some recent experiments we have conducted in our laboratory that have examined the role that informational masking plays in speech perception and attempted to isolate the effects that informational and/or energetic masking have on multitalker listening.

Methods:

All of the experiments described in this chapter were conducted using the Coordinate Response Measure (CRM). In the CRM task, a listener hears one or more simultaneous phrases of the form "Ready, (Call Sign), go to (color) (number) now" with one of eight call signs ("Baron," "Charlie," "Ringo," "Eagle," "Arrow," "Hopper," "Tiger," and "Laker"), one of four colors (red, blue, green, white), and one of eight numbers (1-8). Researchers at the Air Force Research Laboratory have made a corpus of CRM speech materials available to the public on CD-ROM (Bolia et al., 2000). This corpus contains all 256 possible CRM phrases (8 call signs X 4 colors X 8 numbers) spoken by each of eight different talkers (four male, four female). In the experiments described here, the stimulus always consisted of a combination of a target phrase, which was randomly selected from all of the phrases in the corpus with the call sign "Baron," and one or more masking phrases, which were randomly selected from the phrases in the corpus with different call signs, colors, and numbers than the target phrase. The listener's task was to listen for the phrase containing the pre-assigned target call sign "baron" and respond with the color and number combination contained in that phrase. These stimuli were presented over headphones at a comfortable listening level (approximately 70 dB SPL), and the listener's responses were collected either by using the computer mouse to select the appropriately colored number from a matrix of colored numbers on the CRT or by pressing an appropriately marked key on a standard computer keyboard.

Factors that influence informational and energetic masking in speech perception:
Figure 1 shows performance in the CRM listening task with five different maskers: speech-spectrum-shaped noise that has been amplitude modulated to match the intensity fluctuations that occur in normal speech (TM); continuous speech-spectrum-shaped noise (TN); and a different-sex, same-sex, and same-talker speech signal (TD, TS and TT, respectively). The results shown in this figure highlight three important characteristics of informational masking in speech perception:

1. *The difference between speech-in-noise and speech-on-speech masking:* The two noise conditions shown in Figure 1 (TM and TN) are fundamentally different from the speech conditions in two important ways. First, performance with the noise maskers tends to remain at a high level at much lower SNR levels than performance with the speech maskers. Second, once the SNR does become low enough to degrade

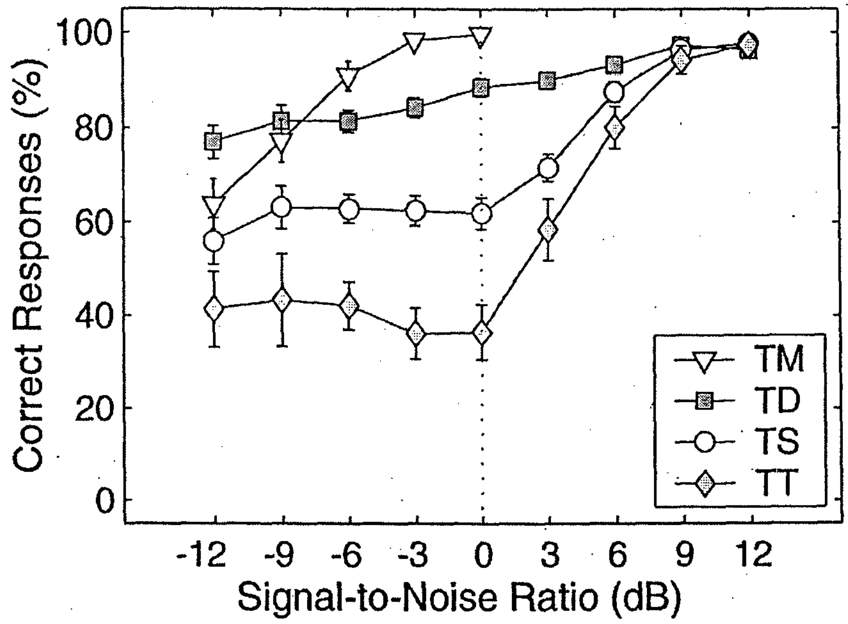

Figure 1: Color and number identifications as a function of signal-to-noise ratio for five type of masking signals: TM- envelope-modulated speech-shaped noise; TN- continuous speech-shaped noise; TD- a different-sex masking phrase from the CRM corpus; TS- a same-sex masking phrase from the CRM corpus; and TT- a masking phrase from the CRM corpus spoken by the same talker used in the target phrase. Adapted from Brungart (2001).

performance with the noise maskers, performance degrades monotonically and precipitously as the SNR is further reduced. In contrast, performance with the speech maskers (TD, TS, and TM) starts to degrade at much higher SNRs but degrades more gradually, especially at negative SNR values.

2. *The importance of voice characteristics:* Performance in the CRM task is much better with a different-sex interfering talker (TD) than with a same-sex interfering talker (TS), and much better with a same-sex interfering talker than with a masking phrase spoken by the same talker used in the target phrase (TT). Because informational masking depends on the relative similarity of the target and masking voices, differences in voice characteristics can be a powerful cue for segregating the target and masking speech signals.

3. *The advantage of level differences:* In contrast to performance with a noise masker, which degrades monotonically as the SNR decreases, performance with a same-sex speech masker tends to plateau around 0 dB SNR. The reason for this plateau in performance is that listeners are able to use differences in the levels of the two talkers to distinguish the two competing voices and selectively attend to the quieter of the two talkers in the stimulus. Thus, especially in the same talker (TT) condition, listeners may do better at negative SNR values because they can identify the target as the quieter talker in the stimulus. In contrast, when the SNR is 0 dB in the TT condition, the prosodic and coarticulative features that connect the call sign and color and number combination in the target phrase are the only available features to allow the listeners to discriminate between the color and number coordinates in target and masking voices.

Figure 2 shows how performance in the CRM listening task changes as additional masking talkers are added to the stimulus. When no competing talkers were present in the stimulus, performance was near 100%. The first competing talker reduced performance by a factor of approximately 0.4, to 62% correct

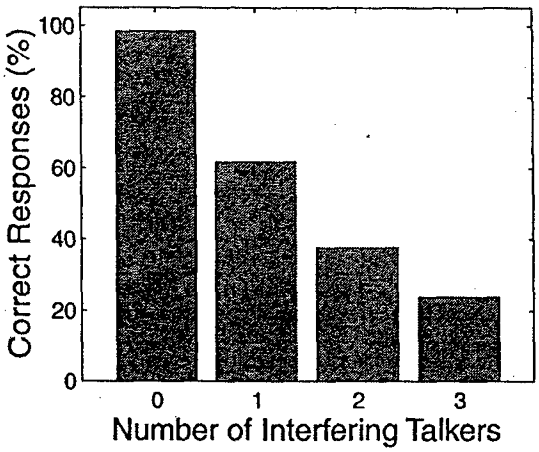

Figure 2: Performance in a diotic CRM listening task with 0, 1, 2, or 3 interfering same-sex talkers.

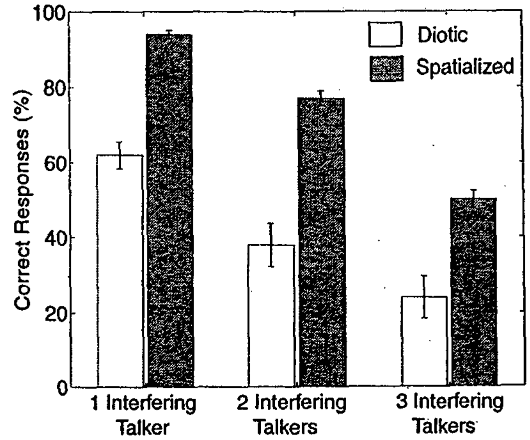

Figure 3: Performance in a CRM listening task with 0, 1, 2, or 3 interfering same-sex talkers, presented diotically or spatially separated by 45 degrees.

responses. The second competing talker reduced performance by another factor of 0.4, to 38% correct responses. And the third competing talker reduced performance by another factor of 0.4, to 24% correct responses. Thus we see that CRM performance in a diotic multitalker speech display decreases by approximately 40% for each additional same-sex talker added to the stimulus.

In general, informational masking is reduced whenever the attributes of the competing talkers are made more distinct in one or more perceptual dimensions. One very powerful way to distinguish the competing talkers in a multitalker stimulus is to spatially separate the apparent locations of the competing talkers. Figure 3 shows performance in the CRM task with 1, 2, or 3 competing talkers both in the diotic condition, where the talkers were presented from the same location, and in a spatial condition, where the talkers were spatially separated by 45 degrees in azimuth. In the case with one interfering talker, spatial separation increased performance by approximately 25 percentage points. In the cases with two or three interfering talkers, spatial separation nearly doubled the percentage of correct responses. These results clearly illustrate the substantial decreases in informational masking that spatial separation in azimuth can produce in multitalker listening.

Figure 4 shows a final example of purely informational masking in dichotic speech perception. In this experiment, the normal two-talker same sex (TS) CRM speech stimulus was presented to the right ear. However, in this case, an additional speech noise masker was presented to the left ear (as indicated in the legend). The listeners were instructed to ignore the left ear and focus only on the right ear. The results show that a speech signal in the left ear interfered substantially with performance even when it was presented at a level 15 dB below the level of the target talker in the right ear, but that a noise signal in the left ear did not interfere even when it was presented at a level 20 dB louder than the target speech signal. In this case, the interference that occurred in the contralateral speech conditions was purely informational and had no energetic component. Ongoing research in our laboratory is now attempting to find other ways to isolate the informational and energetic components of speech on speech masking. Our hope is that this will result in a more complete understanding of the informational masking that occurs in speech and, in the long term, a significant improvement both in the audio displays that are used for multichannel speech communications and in the ability of automatic speech processing systems to process multitalker speech signals.

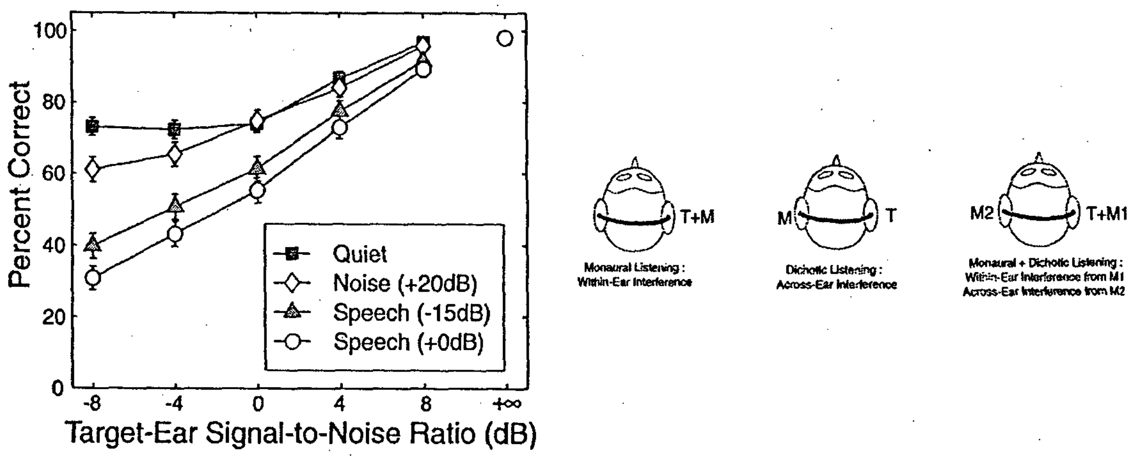

Figure 4: Performance in a dichotic CRM listening task with the target and one same-sex talker presented in the right ear and a masking signal (indicated by the legend) presented in the left ear.

References:

Brungart, D.S. and Simpson, B.D. (2003) Within-ear and across-ear interference in a cocktail party listening task: effects of masker uncertainty. In press, *Journal of the Acoustical Society of America.*

Darwin, C.J., Brungart, D.S. and Simpson, B.D. (2003) Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America* 114(5), 2913-2922.

Kidd, G, Mason, C.R., Arbogast, T.L., Brungart, D.S. and Simpson, B.D. (2003) Informational masking caused by contralateral stimulation. *Journal of the Acoustical Society of America* 113(3) 1594-1603.

Ericson, M.A. and Brungart, D.S. and Simpson, B.D. (2003) Factors that influence intelligibility in multitalker speech displays. In press, *International Journal of Aviation Psychology.*

Brungart, D.S. and Simpson, B.D. (2002) Within-ear and across-ear interference in a cocktail-party listening task. *Journal of the Acoustical Society of America* 112(6) 2985-2905.

Brungart, D.S. and Simpson, B.D. (2002) The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. *Journal of the Acoustical Society of America*, 112(2), 664-676.

Brungart, D.S., Simpson, B.D., Scott, K.R. and Ericson, M.A. (2001) Informational and energetic masking effects in the perception of multiple simultaneous talkers. *Journal of the Acoustical Society of America*, 110(5), 2527-2538.

Brungart, D.S. (2001) Evaluation of speech intelligibility with the Coordinate Response Measure. *Journal of the Acoustical Society of America.* 109(5), 2276-2279

Brungart, D.S. (2001) Informational and energetic masking effects in the perception of two simultaneous talkers. *Journal of the Acoustical Society of America*, 109(3), 1101-1109.